



## 【特許請求の範囲】

【請求項1】 音声の標本化データを一定時間集積したフレームデータの示す音声进行分析した分析結果を出力する音声情報分析方法であって、

各フレームデータの表す音声にノイズ以外の音声が含まれているか否かを判定するステップと、

ノイズ以外の音声が含まれていないと判定された場合に、フレームデータからフレームデータの表す音声に含まれているノイズの特徴を抽出して記憶し、前記フレームデータの示す音声の分析結果として、あらかじめ用意した無音の音声の分析結果を表す情報を出力する無音系処理を実行するステップと、

ノイズ以外の音声が含まれていると判定した場合に、前回の無音系処理で記憶したノイズの特徴分をフレームデータの表す音声から除去し、ノイズの特徴分を除去したフレームデータの示す音声进行分析し、分析した結果を出力する有音系処理を実行するステップとを有することを特徴とする音声情報分析方法。

【請求項2】 音声の標本化データを一定時間集積したフレームデータの示す音声进行分析した分析結果を出力する音声情報分析装置であって、

ノイズテーブルと、

各フレームデータの表す音声にノイズ以外の音声が含まれているか否かを、フレームデータの表す音声の振幅に基づいて判定する手段と、

ノイズ以外の音声が含まれていないと判定された場合に、フレームデータからフレームデータの表す音声に含まれているノイズの特徴を抽出して前記ノイズテーブルに記憶し、前記フレームデータの示す音声の分析結果として、あらかじめ用意した無音の音声の分析結果を表す情報を出力する手段と、

ノイズ以外の音声が含まれていると判定された場合に、前記ノイズテーブルに記憶されているノイズの特徴分をフレームデータの表す音声から除去し、ノイズの特徴分を除去したフレームデータの示す音声进行分析し、分析結果を出力する手段とを有することを特徴とする音声情報分析装置。

【請求項3】 音声の標本化データを一定時間集積したフレームデータの示す音声进行分析し、各フレームデータの表す音声のピッチを表すピッチ情報と、前記ピッチ内の音声波形を表す波形情報と、フレームデータの表す音声の振幅を表す振幅情報とを出力する音声情報分析装置であって、

各フレームデータの表す音声のパワースペクトルを求める手段と、

求められたパワースペクトルからフレームデータの示す音声のピッチ情報を抽出して出力する手段と、

ノイズテーブルと、

各フレームデータの表す音声にノイズ以外の音声が含まれているか否かを、フレームデータの表す音声の振幅に

基づいて判定する手段と、

ノイズ以外の音声が含まれていないと判定された場合に、当該フレームデータの表す音声のパワースペクトルからフレームデータの表す音声に含まれているノイズの特徴を抽出して前記ノイズテーブルに記憶する手段と、ノイズ以外の音声が含まれていないと判定された場合に、前記フレームデータの示す音声の分析結果として、あらかじめ用意した無音の波形を表す波形情報を出力する手段と、

10 ノイズ以外の音声が含まれていると判定された場合に、前記ノイズテーブル記憶されているノイズの特徴分をフレームデータの表す音声のパワースペクトルから除去する手段と、ノイズ以外の音声が含まれていると判定された場合に、ノイズの特徴分を除去したパワースペクトルから前記波形情報と振幅情報を、対応するフレームデータについて求められた前記ピッチ情報を用いて抽出して出力する手段とを有することを有することを特徴とする音声情報分析装置。

【請求項4】 音声の標本化データを一定時間集積したフレームデータの示す音声进行分析し、各フレームデータの表す音声のピッチを表すピッチ情報と、前記ピッチ内の音声波形を表す波形情報と、フレームデータの表す音声の振幅を表す振幅情報とを出力する音声情報分析装置であって、

各フレームデータの表す音声のパワースペクトルを求める手段と、

求められたパワースペクトルからフレームデータの示す音声のピッチ情報を抽出して出力する手段と、

ノイズテーブルと、

30 各フレームデータの表す音声にノイズ以外の音声が含まれているか否かを、当該フレームデータについて求められたピッチ情報、または、当該フレームデータについて求められたピッチ情報と当該フレームデータの表す音声の振幅に基づいて判定する手段と、

ノイズ以外の音声が含まれていないと判定された場合に、求められたパワースペクトルからフレームデータの表す音声に含まれているノイズの特徴を抽出して前記ノイズテーブルに記憶するノイズ処理手段と、ノイズ以外の音声が含まれていないと判定された場合に、前記フレームデータの示す音声の分析結果として、あらかじめ用意した無音の波形を表す波形情報を出力する手段と、

40 ノイズ以外の音声が含まれていると判定された場合に、前記ノイズテーブル記憶されているノイズの特徴分をフレームデータの表す音声のパワースペクトルから除去する手段と、ノイズ以外の音声が含まれていると判定された場合に、ノイズの特徴分を除去したパワースペクトルから前記波形情報と振幅情報を、対応するフレームデータについて求められた前記ピッチ情報を用いて抽出して出力する手段とを有することを有することを特徴とする音声情報分析装置。

【請求項5】請求項3または4記載の音声情報分析装置と、前記音声情報分析装置が出力する波形情報を量子化し、量子化データを出力する手段とを有することを特徴とする音声圧縮符号化装置。

【請求項6】音声を入力する手段と、入力された音声を標本化し標本化データを出力する手段と、標本化データを集積し、標本化データを一定時間集積した前記フレームデータ出力するバッファ手段と、請求項5記載の前記音声圧縮符号化装置と、前記音声圧縮符号化装置の出力する前記ピッチ情報と波形情報と振幅情報を、有線もしくは無線の伝送路を介して送信する手段とを有することを特徴とする通信端末装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、音声情報分析を行う音声情報分析装置に関し、特に、音声信号から動的に変化するノイズ成分を除去する技術に関するものである。

【0002】

【従来の技術】入力音声信号からノイズを除去する従来の技術としては、特開平2-278298号公報記載の技術や、特開平1-75593号公報記載の技術が知られている。

【0003】前記特開平2-278298号公報記載の技術は、フィルタを細分化して使用し、固定的に特定の周波数の除去する技術であり、その音質は除去する周波数によって左右される。また、前記特開平1-75593号公報記載の技術は、ノイズ除去用のニューラル・ネットワークを用いて、あらかじめノイズの特徴を学習／抽出し、これを用いてノイズを除去する技術であり、その音質は、ノイズの特徴を学習／抽出に用いるデータ（音声＋ノイズ）に混入するノイズ選定に左右される。

【0004】

【発明が解決しようとする課題】前記、特開平2-278298号公報記載のノイズ除去用のフィルタを用いて特定の周波数を除去する技術によれば、除去対象とする周波数のが固定的であるため、話者、背景の移動、変化に伴い動的に特徴が変化するノイズの除去には適していない。

【0005】また、前記特開平1-75593号公報記載のノイズの特徴を、あらかじめ抽出／学習させておく技術でも、ニューラル・ネットワークの性質上、ノイズ学習には長時間を要するので、話者、背景の移動、変化に伴い動的に特徴が変化するノイズの除去には適していない。

【0006】また、音声情報分析を行う場合に、音声情報分析対象の音声信号から、ノイズを除去するために、これらの技術を適用すると、いずれの場合も、音声情報分析処理の前処理として、声分析処理とは別個にノイズ除去の処理を行わなければならない。このため、リアルタイムな音声情報分析処理を行う場合には、過負荷とな

る可能性があり、音声情報分析の機能を制限して実現するか、各処理を、異なるプロセッサで実行させることにより実現する必要が生じる場合がある。

【0007】そこで、本発明は、音声情報分析の対象とする音声信号から、動的に特徴が変化するノイズの除去を効率良く行うことのできる音声情報分析装置を提供することを目的とする。

【0008】

【課題を解決するための手段】前記目的達成のために、本発明は、音声の標本化データを一定時間集積したフレームデータの示す音声を分析した分析結果を出力する音声情報分析方法であって、各フレームデータの表す音声にノイズ以外の音声が含まれているか否かを判定するステップと、ノイズ以外の音声が含まれていないと判定された場合に、フレームデータからフレームデータの表す音声に含まれているノイズの特徴を抽出して記憶し、前記フレームデータの示す音声の分析結果として、あらかじめ用意した無音の音声の分析結果を表す情報を出力する無音系処理を実行ステップと、ノイズ以外の音声が含まれていると判定した場合には、前回の無音系処理で記憶したノイズの特徴分をフレームデータの表す音声から除去し、ノイズの特徴分を除去したフレームデータの示す音声を分析し、分析した結果を出力する有音系処理を実行するステップとを有することを特徴とする音声情報分析方法を提供する。

【0009】

【作用】本発明に係る音声情報分析方法によれば、各フレームデータの表す音声にノイズ以外の音声が含まれているか否かを判定し、ノイズ以外の音声が含まれていないと判定された場合に、フレームデータからフレームデータの表す音声に含まれているノイズの特徴を抽出して記憶すると共に、前記フレームデータの示す音声の分析結果を表す情報を出力する無音系処理を実行する。一方、ノイズ以外の音声が含まれていると判定した場合には、前回の無音系処理で記憶したノイズの特徴分をフレームデータの表す音声から除去し、ノイズの特徴分を除去したフレームデータの示す音声を分析し、分析した結果を出力する有音系処理を実行する。

【0010】したがって、ノイズ以外の音声が含まれていない期間、すなわち無音とみなして、あらかじめ求める無音の分析結果を出力すれば足りる期間に、常に最新のノイズの特徴抽出を行っておくので、ノイズ以外の音声が含まれている期間には、この抽出した最新のノイズの特徴を用いて、ノイズの除去を行うことができる。また、有音系処理と無音系処理は、同時に生起することは無いので、この処理の実行負荷は小さく、音声情報分析の機能を制限せずに単一のプロセッサ上で実現できる。

【0011】

【実施例】以下、本発明の一実施例を説明する。

【0012】まず、第1の実施例について説明する。

【0013】図5に、本発明に係る音声情報分析装置を適用した通信システムの構成を示す。

【0014】図中、2000が送信装置、1000が受信装置である。

【0015】送信装置2000は、音声信号を音声分析を利用した手法で圧縮符号化して得られたレベル情報と量子化データとピッチ情報を受信装置1000に送信する。受信装置1000は受信した情報より音声を復号化して出力する。

【0016】ここで、前記送信装置200は、送信部900、ベクトル量子化部800、音声情報分析装置100とを備えており、受信装置1000は、受信部1100と、ベクトル逆量子化部1100と、合成部1200と、D/A変換部1300と、バッファメモリ1500と、音声出力装置1400を備えている。

【0017】送信装置2000において、音声情報分析装置100は、入力された入力音声进行分析し、得られたレベル情報681と、ピッチ情報651を送信部900に、正規化波形系列682をベクトル量子化部800に送る。ベクトル量子化部800は受け取った正規化波形系列682をベクトルコードに変換し、変換して得られた量子化データ801を送信部900に送る。送信部900は、レベル情報681と、ピッチ情報651と、量子化データ801を受け取り、これを有線/無線を介して、受信装置1000に送信する。

【0018】一方、受信装置1000において、受信部1000は、送信装置2000より送信された情報を受信し、レベル情報681'と量子化データ801'とピッチ情報651'とを出力する。ベクトル逆量子化部1100は、出力された量子化データ801'を逆量子化する。合成部1200は、ベクトル逆量子化部1100が出力する波形682'を、前記ピッチ情報651'に基づき、繰り返し周期(ピッチ情報)毎に重ね合わせる\*

次フレームデータ閾値 =  $\omega$  前フレームデータの閾値

+ (1 -  $\omega$ ) 現無音パワー値. . . (式1)

すなわち、無音と判定されたときに、無音と判定されたパワーの値と、前フレームデータの判定用いた閾値とに、適当な重み付けを行い、これの加算値を次フレームデータの判定に用いる閾値を求めるようにする。

【0024】また、音声パワー判定部510の有音/無音の判定は、次のように行うようにしても良い。すなわち、判定対象フレームデータのパワーと前フレームデータのパワーとを比較し、差が所定値より大きい場合には、有音/無音の状態が前回と変化したと判定し、差が所定値より小さい場合には、有音/無音の状態が変化していないと判定する。そして、記憶しておいた前フレームデータの判定結果より、判定対象フレームデータについての有音/無音の判定を行う。または、差が所定値より大きい場合に、さらに、判定対象フレームデータのバ

\*ことにより、波形を合成し、バッファメモリ1500格納する。D/A変換部1300は、バッファメモリ400の出力をデジタル/アナログ(D/A)変換する。音声出力装置1400は、D/A変換器1300により得られた音声を出力する。

【0019】以下、前記音声情報分析装置100の詳細について説明する。

【0020】図5に示すように、前記音声情報分析装置100は、音声の入力手段である音声入力部200と、入力された入力音声をアナログ/デジタル(A/D)変換して、音声標本化データに変換するA/D変換器300と、この音声標本化データを、順次記憶するバッファメモリ400を有している。バッファメモリ400に記憶された一定時間(10~30ミリ秒)毎の音声標本化データは、一定時間音声標本化データ(フレームデータ)401として有音/無音判定部500に送られる。また、音声情報分析装置100は、有音/無音を判定する有音/無音判定部500と、音声情報分析装置100は、フレームデータ401から、正規化波形系列682、レベル情報681と、ピッチ情報651を作成する分析部600とを有している。

【0021】まず、有音/無音判定部500は、図1に示すように、音声パワー判定部510から構成される。音声パワー判定部510は、フレームデータ401の各要素の総和(パワー)を求め、これと、閾値(固定値)と比較した後、閾値より小さければ無音、大きければ有音と判定し、有音/無音を表す有音/無音判定子511を出力する。

【0022】なお、上記閾値で十分な効果が得られなければ、閾値を、式1に従い可変とするようにしてもよい。なお、式1において、 $\omega$ は、 $0 < \omega < 1$ を満たす重み値である。

【0023】

ワーと前記閾値との比較を行い有音/無音の判定を行うようにする。

【0025】次に、分析部600は、図2に示すように、フレームデータ401からFFT(Fast Fourier Transform: 高速フーリエ変換)によりフレームデータの周波数特性を得るFFT部620と、FFT部620にデータの設定を行うFFTデータ設定部610とを有している。また、FFT部620のFFTにより得られた複素数の絶対値の二乗値、すなわち、パワースペクトルを出力するパワースペクトル変換部630と、パワースペクトルの縦軸をパワースペクトル軸から振幅軸に変換する軸変換部640とを有している。なお、FFT(Fast Fourier Transform: 高速フーリエ変換)は、信号のサンプ

リング値から元波形を周波数と振幅によって再生する技法DFT (Discrete Fourier Transform: 離散フーリエ変換) を更に高速に実現したものであり、このようなFFTによる信号処理技術は、「信号処理入門」雨宮 好文/佐藤 幸男著 オーム社 P106~ 6.3 「高速フーリエ変換」等に詳細に説明されている。

【0026】また、分析部600は、FFT部730とピッチ抽出部650とノイズ除去部660と逆FFT部670と正規化部680とより構成される有音処理系と、ノイズ処理部690と無音データ設定部700とより構成される無音処理系と、ノイズテーブル710と、有音/無音部720とを有している。

【0027】有音/無音部720は、有音/無音判定部500から出力された有音/無音判定子511を判定し、有音の場合は有音処理系に処理を行わせ、無音の場合には無音処理系に処理を行わせる。

【0028】まず、有音/無音部720が無音と判定した場合の無音処理系の動作を説明する。

【0029】この場合、ノイズ処理部690は、軸変換部640の出力より、ノイズの特徴、ノイズテーブル710に記憶する。すなわち、たとえば、軸変換部より軸変換された周波数特性情報、パワースペクトル列をノイズテーブル710に記憶する。無音用フレーム設定部700は、正規化波形系列682として、無音を正規化波形系列表すデータ682bをベクトル量子化部800に出力する。

【0030】次に、有音/無音部720が有音と判定した場合の有音処理系の動作を説明する。

【0031】この場合において、FFT部730は、パワースペクトルの対数値(ケプストラム)からピッチ周期を求めるために用いられる。ピッチ抽出部650は、FFT部730の出力より、音声の特徴(高さ)と繰り返し周期(ピッチ情報)を抽出する。ノイズ除去部660は、ノイズテーブル710に記憶されているノイズの特徴を用いて、軸変換部640の出力する有音時のパワースペクトル列よりノイズを除去する。逆FFT部670は、ノイズの除去されたパワースペクトルを逆FFTする。正規化部680は、ピッチ抽出部から受け取ったピッチ情報を用いて逆FFT部670の出力を正規化し、逆FFTされた結果の最大値を"1"にし、ピッチ内の波形を表す正規化波形系列682aを列正規化波形系列682として出力する。また、逆FFT部670の出力のレベルの情報をレベル情報681として出力する。

【0032】さて、ノイズ除去部660によるノイズ除去は、たとえば次のようにして行う。すなわち、ノイズ除去部660において、ノイズテーブル710の記憶内容に応じて、各周波数に対する0.0~1.0の重み付けを記憶したノイズ・マスクテーブルを作成し、有音時

のパワースペクトル列に対して対応する重みを乗算する。この重みは、ノイズテーブル710に記憶されている無音時のパワースペクトル(ノイズのパワースペクトル)の絶対値の大きい周波数がより小さくなるよう、順に0.0~1.0の重み付けを行う。つまりノイズが顕著に表れている周波数には、1.0以下の値が乗算されるようにすることによって、有音時のパワースペクトルの該当パワースペクトル値を元の値より減少の方向に移行させ、ノイズ・スペクトルが現れない周波数には1.0の重みを付けることによってその周波数のパワースペクトル値がそのまま残す。この結果として、ノイズの取り除かれたパワースペクトル列を得ることができる。

【0033】以下、本発明の第2の実施例について説明する。

【0034】本第2実施例は、前記第1実施例と有音/無音判定部500と分析部600の構成のみが異なる。

【0035】本第2実施例に係る有音/無音判定部500は、図3に示すように、フレームデータをFFT処理するためのデータ設定を行うFFTデータ設定部610と、フレームデータをFFTする手段であるFFT部620と、得られた複素数の二乗和を求めるパワースペクトル変換部630と、上記フレームデータよりフレームデータの総和を取り、閾値と比較し、有音/無音判定子721を出力する音声パワー判定部510と、前記有音/無音判定子の判定をする有音/無音判定部720と、パワースペクトルをFFT処理するためのFFT部620と、上記パワースペクトルの対数、ケプストラムよりピッチ情報を抽出し、無音時にはピッチ周期が一定に定まらないこと(「デジタル信号処理」古井 貞熙著 P57~P59 4.9 ピッチ抽出)を利用して有音状態を判定して、有音/無音判定子721を出力するピッチ抽出部650とを有している。

【0036】さて、音声パワー判定部510は、前記第1実施例に係る音声パワー判定部510と同様に、各フレームデータのパワーより有音/無音を判定する。しかし、前述した判定方法によれば、パワーのみによって有音/無音を判定しているために、無音状態を誤って有音と判定してしまう場合がある。そこで、本第2実施例では、音声パワー判定部510が有音と判定した場合に、有音/無音判定部720によって、FFT部730とピッチ抽出部650を起動し、さらにピッチ周期を利用した有音/無音判定を行う。

【0037】すなわち、音声パワー判定部510が有音と判定した場合、FFT部730は、パワースペクトルをFFT処理する。ピッチ抽出部650は、この出力より、パワースペクトルの対数、ケプストラムよりピッチ情報を抽出し、無音時にはピッチ周期が一定に定まらないこと(「デジタル信号処理」古井 貞熙著 P57~P59 4.9 ピッチ抽出を参照)を利用して有音/無音状態を判定して、有音/無音判定子721を出力する。な

お、ピッチ抽出部650の有音／無音の判定は、時間軸上のピッチ情報を周波数軸上のピッチ周期に変換し、パワースペクトルに対し、ピッチ周期ごとに極大値が存在すれば有音、存在しなければ無音として判定することにより行うようにしてもよい。

【0038】次に、本第2実施例に係る分析部600は、図4に示すように、縦軸をパワースペクトル軸から振幅軸に変換する軸変換部640と、ノイズを除去するノイズ除去部660と、逆FFTする手段である逆FFT部670と、逆FFTされた結果の最大値を”1”にするための正規化部680と、有音／無音部920と、ノイズの特徴を抽出するノイズ処理部690と、無音を出力するためのデータを設定する無音用データ設定部700と、ノイズ・テーブル710を有する。各部の個々の動作は、前記第1実施例の対応部と同様である。ただ、本第2実施例では、ノイズ除去部660と逆FFT部670と正規化部680が有音処理系を構成している。そして、有音／無音部920が、ピッチ抽出部650の出力する有音／無音判定子がある場合のみ、この有音処理系に処理を行わせる。一方、ノイズ処理部690と無音用データ設定部700とより構成される無音処理系の処理は、有音／無音部920と、有音／無音判定部500の有音／無音部720との、少なくともいづれか一方が無音を判定した場合に行われる。

【0039】以上のように、本実施例によれば、絶えず変化する周囲のノイズを常に抽出し、そのノイズ特徴を取り除くことによって、如何なる状況でも的確なノイズ除去を行うことができる。

【0040】さて、前記第1実施例および第2実施例に係る有音／無音判定部500と分析部600の各部の行う処理は、プロセッサ上で動作するプログラムとして実現することができる。この場合、無音時には、無音処理系の処理のみを行い、有音時には有音処理系の処理のみを行えばよく、かつ、無音処理系の前処理と有音処理系の前処理は共通しているので、ノイズ除去を音声分析の前処理として行う従来の技術に比べ、プロセッサの処理の負荷は小さく、単一のプロセッサ上で動作するプログラムとして実現することができる。

【0041】なお、以上の実施例では、通信システムへの適用を例にとり説明したが、本第1、第2実施例に係る音声情報分析装置は、この他、分析部600の分析結果を用いて、音声認識等の処理を行う装置等、多様な装置に適用することができる。

【0042】

【発明の効果】以上説明してきたように、本発明によれば、音声分析の対象とする音声信号から、動的に特徴が

変化するノイズの除去を効率良く行うことのできる音声分析装置を提供することができる。

【図面の簡単な説明】

【図1】本発明の第1実施例に係る有音／無音判定部の構成を示すブロック図である。

【図2】本発明の第1実施例に係る分析部の構成を示すブロック図である。

【図3】本発明の第2実施例に係る有音／無音判定部の構成を示すブロック図である。

【図4】本発明の第2実施例に係る分析部の構成を示すブロック図である。

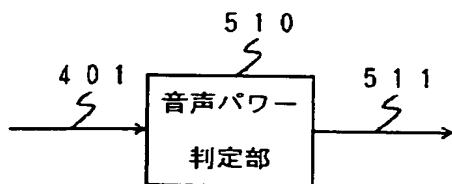
【図5】本発明の実施例に係る通信システムの構成を示すブロック図である。

【符号の説明】

- 100 音声情報分析装置
- 200 音声入力装置
- 300 A/D変換器
- 400 バッファメモリ
- 401 フレームデータ
- 500 有音／無音判定部
- 511 有音／無音判定子
- 600 分析部
- 610 FFTデータ設定部
- 620 FFT部
- 630 パワースペクトル変換部
- 631 パワースペクトル列
- 640 軸変換部
- 650 ピッチ抽出部
- 651 ピッチ情報
- 660 ノイズ除去部
- 670 逆FFT部
- 680 正規化部
- 681 正規化波形系列
- 682 レベル情報
- 690 ノイズ処理部
- 700 無音用フレーム設定部
- 710 ノイズテーブル
- 720 有音／無音部
- 800 ベクトル量子化部
- 801 ベクトルコード
- 900 送信部
- 1100 ベクトル逆量子化部
- 1200 合成部
- 1300 D/A変換部
- 1400 音声出力装置

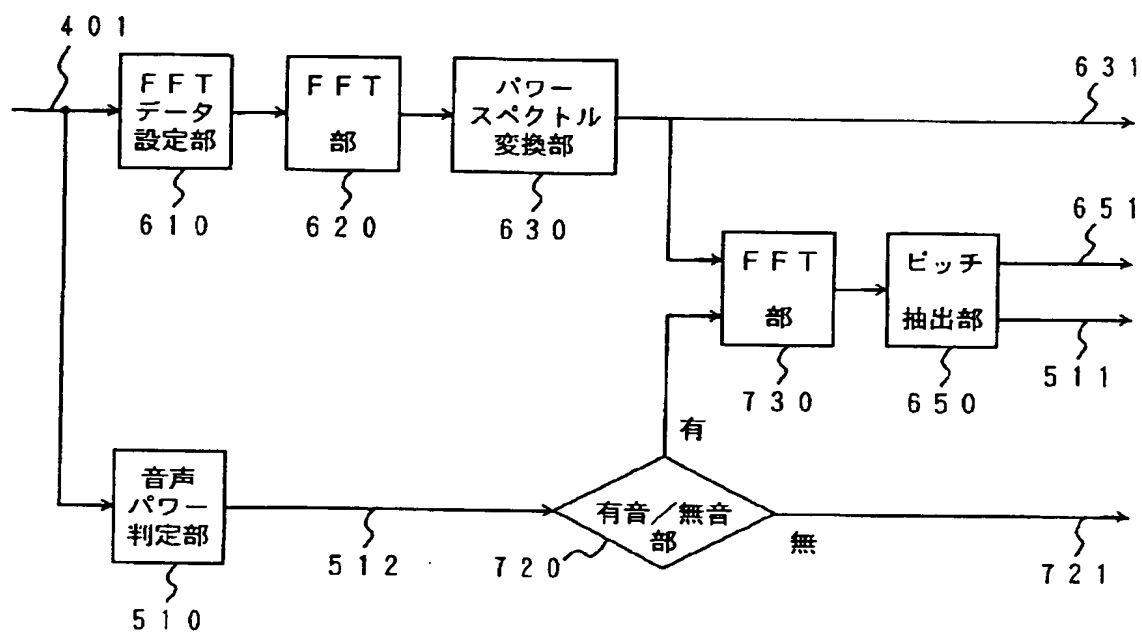
【図1】

図1



【図3】

図3



2





【図4】

図 4

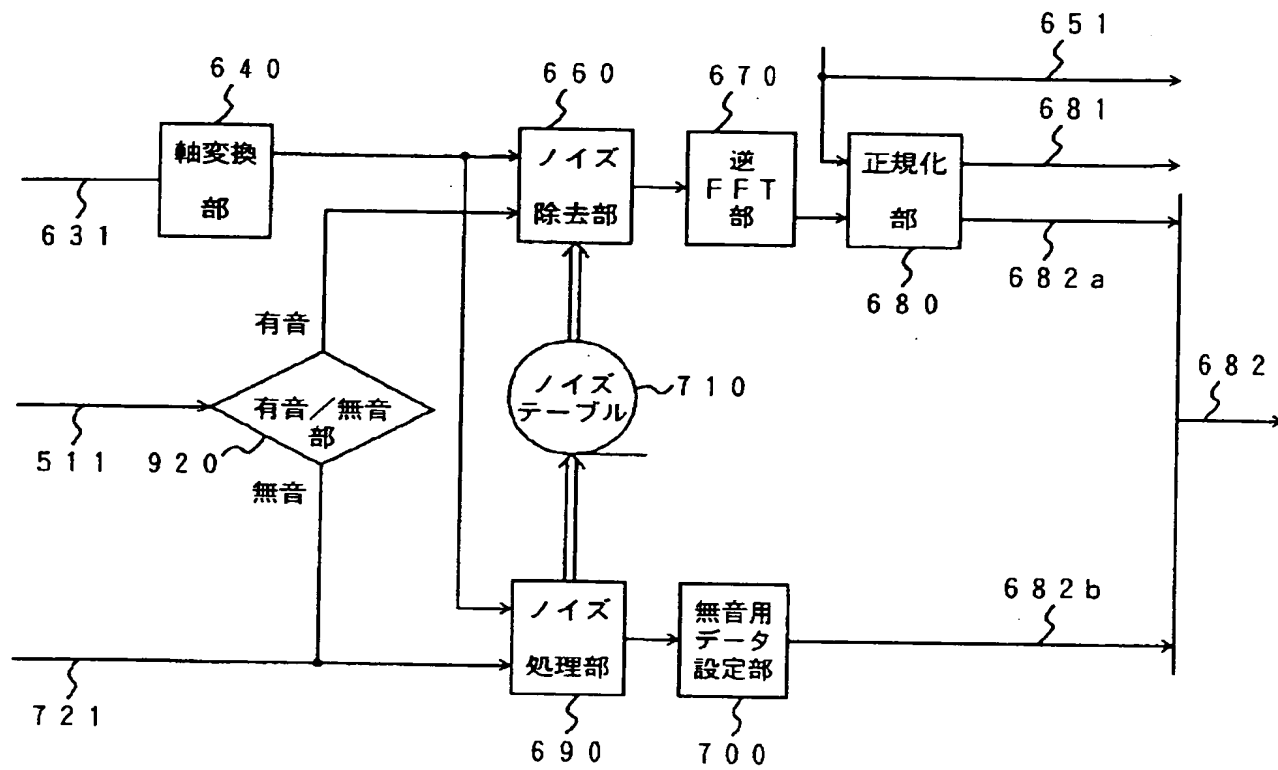


图 5

